Chapter 1

Introduction to Modern Asset Pricing

1.1 A Brief History of Modern Asset Pricing Models

The history of asset pricing is more than three hundred years old but Modern Finance started only about half a century ago. It is generally accepted that Arrow's paper entitled *Optimal Allocation of Securities in Risk Bearing* in 1953 marked the starting point of Modern Finance. In 1874, the French economist, L. Warlas, introduced the concept of general equilibrium, making it the first notion of economic equilibrium in the history of economics study, which was followed by numerous contributions from many famous economists. It was only in 1954 that Arrow and Debreu finally gave a proof to the existence of a general equilibrium.

In Arrow's paper in 1953, he interpreted a financial security as a series of commodities in various future states with different values. This interpretation was later refined by Debreu, who incorporated an equilibrium model over a state space to deal with a financial market so that a security was nothing but a commodity with different values in different states and at different times. His notions of a state price and a state security (or Arrow security) are very popular now. All of these notions are based on the assumption of a complete financial market, that is, corresponding to each contingent state, there is an Arrow security to be traded in a financial market.

When it comes to stock investment, what is its utility? In his doctoral dissertation entitled *Theory of Investment Values* in 1937, William proposed an appraisal model of stock that asks investor to do a long-term forecast of a firm's future dividends and check the accuracy of the forecasting. Later, William described how to combine the forecasting with its accuracy to estimate the intrinsic value of a stock, leading to the well-known DDM (discount dividend model).

When Markowitz studied the DDM model, he found that if all investors follow a DDM model, they would all purchase the stock with the highest expected return and avoid other stocks. This was obviously counter-intuitive. So in his seminal paper in 1952, he proposed that investors need to consider a balance between the return and the risk; indeed he used the mean to describe the expected return and the variance to describe the risk. With this framework, he developed a mean-variance efficient frontier in that given the risk level, an optimal portfolio with the highest expected return is obtained, or equivalently given the expected return level, an optimal portfolio with the minimum risk is obtained. Based on this efficient frontier, the well known two-fund separation theorem was developed to help rational investors to develop optimal investment strategy.

In 1958, Tobin pointed out that when there is a riskless asset, the frontier becomes a straight line and an optimal portfolio is then a combination of a risky asset and the riskless asset. In the 1960s, with all investors having the same expectation, Sharpe, Lintner and Mossin developed the Capital Asset Pricing Model (CAPM) for a financial market at an equilibrium state. One of the important contributions from the CAPM is that it links excess return with the so-called market return. Merton in 1973 developed an Intertemporal Capital Asset Pricing Model, which links the excess return of a risky asset to not only the market return but also several state variables that will eventually result in multi-factors. In 1976, Ross proposed the Arbitrage Pricing Theory: given a financial market spanned by a number of factors, asset pricing of no-arbitrage is based on the results of the factor premiums and factor sensitivities.

In the 1970s, Lucas developed the consumption-based asset pricing model, and in the 1990s all the above models were merged into a more general pricing model, namely the stochastic discount factor (SDF) pricing model, which we discuss in detail as follows.

Modern consumption theory started in the 1930s, when Keynes in his famous book *The General Theory of Employment, Interest, and Money* proposed an Absolute Income Hypothesis based on the Fundamental Psychological Law. Specifically,

- there exists a stable functional relationship between real consumption and income;

- marginal propensity to consumption is bigger than 0 but less than 1;
- average propensity to consumption is decreasing along with increase of income.

Compared with Keynes' absolute income hypothesis, Duesenberry's relative income hypothesis in 1949 was an advancement. Duesenberry stressed the effect of consumption habit. Later, Modiglian and Brumberg (1954) proposed the Life Cycle Hypotheses and Friedman (1957) proposed the Permanent Income Hypotheses. The above work laid the foundation of modern consumption theory.

Modiglian and Brumberg assumed that there is a utility of aggregate consumption depending on historic and future consumption paths. Friedman (1957) divided income into two parts: a predictable income (called a permanent income) and an unpredictable income (called a temporary income). The Permanent Income Hypothesis claims that an individual consumption is not decided by the income of that period but by a life-long income i.e. the permanent income. The mathematical tools used by these authors were deterministic.

In the 1970s, there were two important events that influenced the development of the consumption theory: one is Lucas's (1976) critique on rational expectation and another is Hall's martingale model of consumption (1978). Both stressed expectation and uncertainty. Lucas argued that consumption depends on expected income and Hall proved that consumption follows a martingale process if the preference of the consumer is time-separable, the utility is of a quadratic form and the interest rate is constant. Hall's result shows that the Life Cycle Hypothesis and the Permanent Income Hypothesis follow a random walk. Let c_t denote the consumption at time t and E_t the conditional expectation given the information set up to time t. Then Hall's conclusion is

$$E_t c_{t+1} = c_t, \tag{1.1}$$

or equivalently by using the representation of a martingale difference, a consumption dynamic process follows

$$c_{t+1} = c_t + \eta_{t+1} \tag{1.2}$$

with η_{t+1} being a zero-mean normal white noise. Campbell and Cochrane (2000) stated, 'The development of consumption-based asset pricing theory ranks as one of the major advances in Financial Economics during the last two decades.' This comes from a very intuitive construction dealing with the tradeoff between investment and consumption. Specifically, let $e_{i,t}$ be the endowment of the *i-th* agent at time $t = 0, 1$, $c^j_{i,t}$ be his consumption of the *j-th* commodity (in physical unit or in monetary unit) at time $t = 0, 1$, $j = 1, 2, \cdots$. Let x_j be the payoff of a security which is a financial contract enforced among agents at time $t = 0$; the contract promises to pay back x_j units of commodity j at time $t = 1$. Here, x_j is often a random variable. The action of entering a financial contract is called an investment. Then a tradeoff is attained, for each agent, if endowment $e_{i,0}$ is allocated between the current consumption $c^j_{i,0}$ and the investment $w_{j,i}$, which is the number of contracts (bought or sold) for payoff x_j. People with great patience tend to consume less now and invest more. People with low risk aversion tend to be more involved in highly risky investments, in the hope of obtaining higher level of consumption in the future. The equilibrium allocation for each agent is to maximize his utility, given a budget constraint. Mathematically this is

$$\max_{\{w_{j,i}\}} \{E[u_i(c_0, c_1)]\} \tag{1.3}$$

with the budget constraint

$$c_0 = \sum_j c^j_{i,0} = e_{i,0} - \sum_j w_{i,j} p_j, \tag{1.4}$$

$$c_1 = \sum_j c^j_{i,1} = e_{i,1} + \sum_j w_{i,j} x_j, \tag{1.5}$$

where p_j is the price of the financial contract for commodity j at time $t = 0$, $u_i = u_i(c_0, c_1)$ is a utility function for agent i, and E is the expectation of a random variable. By a simple calculation, we derive an equation of the first condition for utility maximization - the so-called Euler equation:

$$p_j = E[IMRS_i x_j],\ j = 1, 2, \cdots, \tag{1.6}$$

where $IMRS_i$ is the inter-temporal marginal rate of substitution given by $IMRS_i = \frac{\partial u_i/\partial c_0}{\partial u_i/\partial c_1}$, where $\partial u_i/\partial c_i,\ i = 0, 1$, denotes the partial derivative of the utility with respect to the consumptions. The pioneers Lucas (1978), Breeden (1979), Grossman and Shiller (1981), and Hansen and Singleton (1982, 1983) studied the Euler equation in two ways: (i) To determine the assets' prices given the agent's utility function and the assets' payoffs; (ii) To determine, as an inverse problem, the parameters such as the risk aversion in

$IMRS$ by the GMM estimation procedure in Hansen and Singleton (1992, 1983), given the market prices and asset payoffs.

In recent years, people have started to consider a more general form of the Euler equation by introducing a random variable m satisfying

$$p = E[mx]. \tag{1.7}$$

Then m is called a stochastic discount factor (SDF). This framework includes the consumption-based asset models, the CAPM, the Ross arbitrage pricing theorem, the option pricing models and many other popular asset pricing models as special cases. For further details, the readers may wish to consult Cochrane (2001).

Alternatively m is called a state price density (SPD), which is a very popular name in the risk-neutral pricing world.

1.2 The Equity Premium Puzzle

The so-called risk premium is a monetary cost of uncertainty. Its general definition is as follows.

Definition 1.2.1 *Suppose we are given an initial wealth w_0 and investment $l = \langle p_1, x_1; p_2, x_2; \cdots; p_n, x_n \rangle$, in which the investment has an outcome x_i with probability p_i, $i = 1, 2, \cdots, n$; here x_i could be money, a commodity or some other type. Let $w = w_0 + x$ be the terminal wealth. Let r be a real number satisfying*

$$u(Ew - r) = E[u(w)], \tag{1.8}$$

where $u(\cdot)$ is a utility function, E is the expectation, and $x = \{x_1, x_2, \cdots, x_n\}$. Then r is called the risk premium.

An intuitive interpretation of the risk premium is that an investor with utility u is indifferent to risky investment w and a deterministic investment (such as bank's deposit) with a fixed payoff $Ew - r$. To a risk averse investor, we expect that $r > 0$. By the no arbitrage principle, we can see $Ew - r = R_f$, where R_f is a risk-free interest rate. That is, for a risk averse investor to enter the risky trade w, the expected return of w must be higher than the risk-free interest rate by

$$Ew = R_f + r,$$

so r is an expected excess return to compensate the investor for entering the risky trading. When x is an equity, the risk premium becomes an equity premium. It is the difference between the expected equity return and the risk-free interest rate.

Mehra and Prescott (1985) announces a surprising discovery to the financial economics community, which later becomes the well-known *equity premium puzzle*. The puzzle says that there exists a substantial difference between the equity premium estimation by using the U.S. historical stock market data and that by a slight variation of the Lucas model (1978) based on the U.S. aggregate consumption data, unless the risk aversion parameter is raised to an implausibly high level. In the above, we have described briefly a justification for the consumption-based asset pricing modelling. Now, the puzzle represents a serious challenge to the development of Financial Economics. Since its announcement, it has attracted the attention of many financial economists. Specifically, Mehra and Prescott (1985) starts with the problem of seeking an optimal solution for the following maximization problem, which is a variation of the Lucas model (1978) in a pure exchange economy:

$$\max_{c_t} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right], \qquad 0 < \beta < 1, \tag{1.9}$$

subject to the budget constraints above, where β is a subjective discount factor. The utility function is restricted to be of constant relative risk aversion(CRRA) and takes the form

$$u(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma}, \tag{1.10}$$

where γ is the coefficient of relative risk aversion. After some calculations, the following Euler equation emerges

$$p_t = E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} x_{t+1} \right], \tag{1.11}$$

where p_t is the market price at time t, x_{t+1} the stock's payoff at time $t+1$ with $x_{t+1} = p_{t+1} + d_{t+1}$, and d_{t+1} the stock's dividend.

Turning to empirical analysis, the paper finds that the average real return on the riskless short-term securities over 1889-1978 period is 0.8% and the average real return on the S&P 500 composite stock index over the same period is 6.98% per annum. This leads to an average equity premium of 6.18%. By varying γ between 0 and 10, and β between 0 and 1, they find that it is not possible to obtain a matching sizable equity premium. The largest premium from the simulation is only 0.35%! The paper concludes that their model of asset returns is inconsistent with the U.S. data on consumption and asset returns.

Given the market traded prices of some risky assets and the riskfree asset, the aim is to check if an SDF can price them correctly. Instead of checking the Euler equation directly, which is often complex, let m be the SDF that, according to the Euler equation, incurs no pricing error to the risky assets and the riskfree asset. Then Hansen and Jagannathan (1991) shows that there exists a relationship between m and the assets by introducing an upper bound for the so-called Sharpe ratio of the returns of risky assets, namely

$$\frac{|ER - R_f|}{\sigma_R} \leq SR^{max}, \tag{1.12}$$

where R is the (aggregate) return of the risky assets, ER the expectation of R, R_f the return of the riskfree asset, σ_R the volatility of R and SR^{max} the theoretically maximal Sharpe ratio. Here, $SR^{max} = \sigma_m/Em$, where σ_m is the volatility of m and Em the expectation of m. This latter ratio is called the Hansen-Jagannathan bound, or the H-J bound for short. There are two ways to exploit it. First, the absolute value of the Sharpe ratio of the return of risky assets has an upper bound given the SDF. In other words, any risky asset with a Sharpe ratio higher than the SR^{max} cannot be correctly priced (*i.e.* priced with no pricing error) by the SDF m. Alternatively, a necessary condition for an SDF candidate to be capable of pricing the assets correctly is that its SR^{max} must not violate a lower bound which is given by the maximum of the Sharpe ratios of the assets to be priced correctly. Either way the H-J bound provides us with a necessary and quick way to check whether the Euler equation is satisfied.

When the SDF m is given by the CRRA utility, we have $m = \beta(\frac{c_1}{c_0})^{-\gamma}$. Using the yearly U.S. economic and stock data, the inequality (1.12) is overwhelmingly violated, unless the risk aversion parameter γ reaches an extremely high level, say between 25 and 50 or even high. Mathematically, the puzzle can be stated as follow.

Given a series of payoffs $\{x_i\}$ for risky assets and their corresponding market prices $\{p_i\}$, let candidate SDF m be a series of SDFs given by $\{m = m_\theta,\ \theta \in \Theta\}$, where Θ is a parameter space in $\Re^L$. Let Θ_1 be a subspace of Θ whose range is feasible economically. For example, Θ can be defined by $\Theta = \{\gamma\ |0 < \gamma < 1000\}$ and Θ_1 by $\Theta_1 = \{\gamma\ |0 < \gamma < 5\}$. Then a 'parameterized' equity premium puzzle takes the form

$$p_i \neq E[m_\theta x_i] \text{ for some } i \text{ and } \forall\theta \in \Theta_1, \tag{1.13}$$

but

$$p_i = E[m_\theta x_i] \text{ for all } i \text{ and for some } \theta \in \Theta. \tag{1.14}$$

Hansen and Jagannathan (1997) introduces a distance to measure pricing errors for given asset pricing models. The distance becomes the well-known Hansen-Jagannathan distance. Both the H-J bound and the Hansen-Jagannathan distance are powerful ways to evaluate asset pricing models.

Kocherlakota (1996) points out that there are three possible ways to resolve the equity premium puzzle.

- Introduce progressively a more complex form of the utility function, e.g. by some judicious modification of the standard power utility function if time and state are separable, or by adopting a completely new form of the utility function if time and state cannot be separated.

- Traditional asset pricing models make the strong assumption that asset markets are complete, but the real world is not like that. Acknowledging that the market is incomplete in one's asset pricing model is one way to resolve the equity premium puzzle; this will be a new direction.

- Trading costs, such as taxes and brokerage fees, should be included in the asset pricing model because their effects on the equity premium cannot be ignored.

Mehra (2003) points out that the equity premium puzzle is a quantitative puzzle. The puzzle arises from the fact that a quantitative prediction of equity premium is much different from what has been historically documented. The puzzle cannot be dismissed lightly because our economic intuition is often based on our acceptance of models such as CCAPM (Consumption-based CAPM). Therefore, the validity of using such models for any quantitative assessment has also become an issue.

Over the past 20 years, attempts to resolve the puzzle have become a major research activity in Finance and Economics. Many different approaches have been adopted. These include, among many others, the recursive utility model by Epstein and Zin (1989,1991) and Weil (1989,1990), the habit formation model by Constatinides (1990), Abel (1990), and Campbell and Cochrane (1999), the three-state economy by Rietz (1988), the survivorship bias by Brown, Goetzmann, and Ross (1995), the idiosyncratic risk and incomplete markets considered by Mankiw (1986), Lucas (1994) and Telmer (1993), the generalized heterogeneous consumers specified by Constantinides and Duffie (1996) and Brav, Constantinides and Geczy (2002), the market imperfections models by Aiyagari and Gertler (1991), Alvarez and Jermann (2000), Bansal and Coleman (1996), Constantinides, Donaldson and Mehra (2002), Heaton and Lucas (1996), and McGrattan and Prescott (2001). In fact, hundreds of papers have been published in the Finance and Economics literature that are devoted to the puzzle. The puzzle is not only relevant to the foundation of the theory of Financial Economics but also crucial to the financial industry for such activities as the long-term asset allocation for pension funds, whose market size currently exceeds one trillion U.S. dollars. In the following, we give a brief introduction to some of the recent developments.

I. Preference Modifications

I.a Generalized Expected Utility

The standard preference class used in macroeconomics consists of time-and-state-separable utility functions. From the empirical analysis of Mehra and Prescott (1985), we know that the CRRA preference can match the observed equity premium only if the coefficient of relative risk aversion is implausibly large.

Epstein and Zin (1989,1991) present the notion of *generalized expected utility preference* (GEU) that allows independent parameterizations of the coefficient of risk aversion and the elasticity of inter-temporal substitution. The recursive utility, U_t, is given by

$$U_t = [c_t^{1-\rho} + \beta(E_t U_{t+1}^{1-\alpha})^{(1-\rho)/(1-\alpha)}]^{1/(1-\rho)}, \tag{1.15}$$

where α measures the relative risk aversion and $1/\rho$ the elasticity of inter-temporal substitution. In contrast to the historical average risk premium of 6.2% and the largest premium obtained by Mehra and Prescott (1985) of 0.35%, Epstein and Zin's results in 1991 show a low riskfree rate and an average equity premium of roughly 2%. In their words, their GEU specification only partially resolves the puzzle. Weil's model in 1990 is very much like that in Epstein and Zin (1991). It is very interesting to note that the Epstein-Zin model gives a two-factor representation for the premium by

$$E_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} = \frac{\theta}{\psi} cov_t(r_{t+1}, \Delta c_{t+1}) + (1-\theta) cov_t(r_{t+1}, r_{p,t+1}). \qquad (1.16)$$

For detailed derivation, see, e.g., Campbell and Viceira (2002, p. 45). We shall give a new result later in the form of (Theorem (3.3.5)), which shows that this two-factor structure of the Epstein-Zin model is the key to the superior performance of the model over models of the CRRA type.

I.b Habit Formation

Abel (1990) and Constantinides (1990) consider the effects of consumption habit on the decision making of individuals. Constatinides specifies a utility as follows:

$$U(c) = E_t \sum_0^{\infty} \beta^s \frac{(c_{t+s} - \lambda c_{t+s-1})^{1-\alpha}}{1-\alpha}, \quad \lambda > 0, \qquad (1.17)$$

where λ is a parameter that captures the effect (or habit) of past consumption. This preference ordering makes individuals extremely averse to consumption risk even when the risk aversion is small.

Abel (1990) provides another kind of habit formation by defining utility of consumption relative to the average per capita consumption. The idea is that one's utility depends not only on the absolute level of consumption but also on how one is doing relative to others. In contrast to Constantinides (1990), the per capita consumption can be regarded as the 'external' habit formation for each individual. Again we shall give new results later to show why such a setup is to do with a more complex utility that goes back to Kocherlakota's point above.

Campbell and Cochrane (1999) specifies habit formation as external, similar to Abel (1990), and took the possibility of recession as a state variable so that a high equity premium could be generated.

As it turns out the habit formation models have only limited success as far as resolving the equity premium puzzle is concerned because effective risk aversion and prudence become improbably large in these models.

II. Incomplete Markets

Instead of assuming all the agents are homogeneous, Mankiw (1986) argues as follows. There are infinitely many consumers who are identical ex ante, but their consumptions are not the same ex post. The aggregate shocks to consumption are assumed to be not dispersed equally across all individuals but only affect some of them ex post. Under the assumption of incomplete markets, Mankiw shows that representative agent models are not effective as approximations to a complex economy with ex post heterogeneous individuals.

Lucas (1994) proposes a more general model than Mankiw (1986) by assuming undiversified shocks to income and borrowing, and short sale constraints at an infinite time horizon. In her model, individuals cannot insure, ex ante, against future idiosyncratic shocks to their income. She showed that individuals with a bad idiosyncratic shock can effectively self-insure by selling financial assets to individuals with good luck by trading. Hence, idiosyncratic risks to income are largely irrelevant to asset prices with trading even when the borrowing constraints are severe. To resolve the equity premium puzzle requires more than closing forward market for labour income. Telmer (1993) considers much a similar incomplete market model as Lucas (1994) by assuming two heterogeneous consumers with different consumption stream in the economy. His research supports Lucas' conclusion.

Constantinides and Duffie (1996) assumes that consumers are heterogeneous because of uninsurable, persistent and heteroscedastic labour income shocks at each time period. The paper constructs a model in which the Euler equations depend on not only the marginal rate of substitution (MRS) at the aggregate level but also the cross-sectional variance of the individual consumption growth. Continuing the work of Contandinides and Duffie (1996), Brav, Constantides and Geczy (2002) considers the case of asset pricing with heterogeneous agents and limited participation of households in capital markets. Their empirical analysis reveals that relaxation of the assumption of complete consumption insurance is helpful in resolving the equity premium puzzle.

III. Liquidity Premium and Trading Costs

Bansal and Coleman (1996) gives a monetary explanation for the equity premium puzzle. In their model, assets other than money play a key feature by facilitating transactions. Using empirical evidence, the paper claims that half of the equity premium can be captured by their model. Some economists try to explain the equity premium puzzle via transaction costs. For example, McGrattan and Prescott (2001) proposes an explanation based on changes in tax rates. Heaton and Lucas (1996) finds that the differences in trading costs across stocks and bond markets have to be very high in order to resolve the equity premium puzzle with transaction costs. Kocherlakota (1996) even shows that to match the real equity premium, the trading cost have to be implausibly high.

Mehra (2003) gives a comprehensive summary of recent developments; the puzzle remains open.